The Spread Optimization Problem

F. Altarelli,[1,2] A. Braunstein,[1,3,2] L. Dall'Asta,[1,2] and R. Zecchina[1,3,2]

[1] *DISAT and Center for Computational Sciences, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy*
[2] *Collegio Carlo Alberto, Via Real Collegio 30, 10024 Moncalieri, Italy*
[3] *Human Genetics Foundation, Via Nizza 52, 10126 Torino, Italy*

We analyze the irreversible dynamical process corresponding to the linear-threshold model of influence spread over a network, and propose an efficient algorithm to solve the inverse problem, namely that of finding an optimal initial condition that generates to a desired final state of the dynamics. This is a challenging problem which is crucial in many contexts, from systemic risk analysis to the design of viral marketing campaigns. Our approach is based on the cavity-method of statistical physics. We compare our algorithm to standard techniques (based on Monte Carlo and Linear Programming) and show that it has an excellent performance.

INTRODUCTION

Much effort has been devoted over the past forty years to investigate the properties of irreversible dynamical systems, such as their relaxation or the dependence of the equilibrium state on the initial condition. Recently the interest extended also outside physics, with the application of statistical physics methods to problems arising in economics, biology and computer science. One limitation of the approaches taken so far is that they provide ways to explain and predict the observed phenomena, but no instrument to control them. This aspect may be unusual in physics, but it is central in the applied sciences. In engineering for instance, the main goal is that of designing the structure of a system or the initial conditions of a dynamical process in such a way to favor/prevent the occurrence of some phenomenon. It follows that one of the challenges in the interdisciplinary research on complex systems should be that of developing efficient methods and algorithms to *control and optimize* their dynamical behavior. This is particularly important when these systems are large networked structures with many degrees of freedom and complex interaction patterns [1].

The simplest form of dynamics taking place on networks is probably that of irreversible spreading processes, in which nodes "activate" depending on the state of their neighbors. Despite the straightforward mathematical formulation, spreading processes find application in a variety of relevant practical problems. For instance, one could be interested in activating the whole graph at time T using a minimum number of "seeds". This is a typical problem faced in the design of viral marketing campaigns, whose goal is that of targeting those individuals that guarantee the maximum spread of the advertisement throughout a social network [2, 3]. A similar optimization process can be used to identify which nodes are most sensitive for the propagation of failures in a power grid or liquidity shocks in interbank networks, providing a tool to assess and prevent systemic risk [4].

In this Letter, we show that the cavity method and message-passing algorithms can be applied to efficiently solve *spread optimization problems* on large graphs, providing a framework to study also more complex inverse dynamical problems. Our results demonstrate that the performance of the optimization depends on the nature of the collective dynamics and, in particular, on the presence of cooperative effects. Moreover, the dynamical trajectories selected under optimization are characterized by peculiar statistical properties, such as a very slow pace with large fluctuations of the activation times. We conclude showing the relation between our Belief Propagation (BP) based solution of the inverse dynamics and recursive equations recently used to study the dynamics of other irreversible processes on networks.

THE SPREAD OPTIMIZATION PROBLEM

We consider a linear-threshold model (LTM) [2] in which the nodes of a graph $G = \{V, E\}$ can deterministically switch from a quiescent state ($x_i = 0$) to an active one ($x_i = 1$) during a discrete-time dynamics. Each directed edge $(i, j) \in E$ is endowed with a *weight* $w_{ij} \in \mathbb{R}^+$, while the nodes are given a set of *thresholds* $\{\theta_i \in \mathbb{R}^+,\, i \in V\}$. A quiescent node i activates when the total weighted influence of its active neighbors exceeds its threshold, i.e. it follows the discrete-time parallel dynamics

$$x_i^{t+1} = \begin{cases} 1 & \text{if } x_i^t = 1 \text{ or } \sum_{j\in\partial i} w_{ji} x_j^t \geq \theta_i\,, \\ 0 & \text{otherwise}\,. \end{cases} \qquad (1)$$

We shall call "seeds" the nodes active at $t = 0$.

The direct dynamics for this model has been recently solved in the context of the zero-temperature random field Ising model [7]. Here we adopt a different approach allowing us to study an associated optimization problem that focuses on large deviations from the typical case analyzed in these works. We shall highlight later the relations between our approach and the non-optimized case.

Because of irreversibility, a trajectory $\{\mathbf{x}^0, \dots, \mathbf{x}^T\}$ can be fully parametrized by a configuration of activation times $\mathbf{t} = \{t_1, \dots, t_N\}$, with $t_i \in \{0, 1, 2, \dots, T, \infty\}$ representing the time at which x_i becomes equal to 1, and conventionally defining $t_i = \infty$ iff $x_i^t = 0$ for any $t \leq T$, an arbitrarily defined stopping time. Given an assignment of the seeds $S = \{i : t_i = 0\}$, the dynamics is fully determined for $i \notin S$ by the set of relations $t_i = \phi_i(\{t_j\})$ with $j \in \partial i$ and $\phi_i(\{t_j\}) = \min_{\sum_{j\in\partial i} w_{ji}\mathbb{1}[t_j<t]\geq\theta_i} t$ (we conventionally assume the minimum of an empty set to be ∞). Defining $\Psi_i(t_i, \{t_j\}) = \mathbb{1}[t_i = 0] + \mathbb{1}[t_i = \phi_i(\{t_j\})]$, the dynamical constraints can be simply expressed by the equations $\Psi_i(t_i, \{t_j\}) = 1$ for every i. To optimize over the assignment of the seeds, we weight each dynamical trajectory with an energetic term $\mathscr{E}(\mathbf{t}) = \sum_i \mathscr{E}_i(t_i)$ where $\mathscr{E}_i(t_i) = \mu c_i \mathbb{1}[t_i = 0] + \epsilon \mathbb{1}[t_i = \infty]$ and $c_i \in]0, 1]$ is the cost to be selected as a seed. The trade-off between the total cost of the seeds and the final number of active nodes is controlled by the chemical potentials $\mu, \epsilon \in \mathbb{R}$. The resulting optimization problem is computationally hard even to approximate in the worst case [14].

THE BP SOLUTION

In order to eliminate systematic short loops due to the fact that the constraints Ψ_i, Ψ_j share the variables t_i, t_j for $(i, j) \in E$, we employ a dual factor graph in which variable nodes are associated to edges $(i, j) \in E$ and represent the pairs of times (t_i, t_j) while the factor nodes contain three terms: 1) the dynamical constraints, 2) the constraint imposing that t_i is the same for all edges incident on i, and 3) the contribution from i to the energy (see Supplemental Material, henceforth SM). Using this representation, the optimization problem can be analyzed on locally tree-like graphs by means of the cavity method and solved using message-passing algorithms. In fact, the (cavity) marginal probability $H_{i\ell}(t_i, t_\ell)$ that nodes i and ℓ activate at times t_i and t_ℓ in absence of Ψ_ℓ satisfies the following BP equations

$$H_{i\ell}(t_i, t_\ell) \propto e^{-\beta\mathscr{E}_i(t_i)} \sum_{\{t_j\}} \Psi_i(t_i, \{t_j\}) \prod_j H_{ji}(t_j, t_i) \quad (2)$$

with $j \in \partial i \setminus \ell$. Given a solution of (2), the marginal probability of (t_i, t_j) is $P_{ij}(t_i, t_j) \propto H_{ij}(t_i, t_j) H_{ji}(t_j, t_i)$.

Eqs. (2) allow to study the statistics of the dynamical trajectories (e.g. entropies or distribution of activation times). For $\beta \to \infty$, with a proper rescaling of the messages, they give the Max-Sum (MS) algorithm, which can be used to find explicit solutions to the spread maximization problem (see SM). Some kind of approximation is often needed in practice to obtain efficient updates. For small connectivity, continuous weights and thresholds must be discretized, in order to have only a finite spectrum of possible values for the sum $\sum_{j\in\partial i} w_{ji}\mathbb{1}[t_j < t]$. We call Θ the integer value of the discretized threshold. We notice that the BP-messages do not depend on all T^2 combinations of times (t_i, t_ℓ), but only on $(t_i, \text{sign}(t_i - t_\ell + 1))$. It follows that the time complexity per node and per iteration of BP and MS is $O(T\Theta^4|\partial i||\partial i - 1|)$. By exploiting the associativity of the convolution needed to perform the local update rule (see SM), the overall time complexity per iteration can be reduced to scale as $O(T\Theta^4|E|)$. In the case of BP, the convolution can then be performed with Fast Fourier Transform, reducing further the complexity to $O(T\Theta^2|E|)$. While it is clear that increasing the resolution on the weights and thresholds could be computationally very expensive, the computational complexity of message-passing is still very low compared to other optimization algorithms such as exhaustive search and Monte Carlo methods. On the other hand, for large connectivity (e.g. fully connected graphs), continuous weights can be kept, and the update can be simplified by using the Central Limit Theorem.

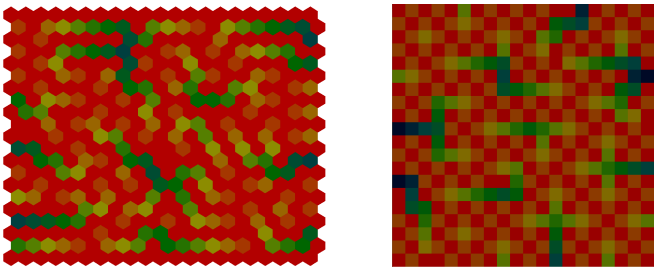

FIG. 1. Example of low energy configurations found by MS equations with reinforcement in 2D $L \times L$ lattices with open boundaries. Colors correspond to activations times (from time zero for seed in red to colder colors). Left: hexagonal lattice ($k = 6$) with $\theta = 5, T = 10, L = 21$. Every hexagon is adjacent to at most one hexagon of a colder color. Right: rectangular lattice ($k = 4$) with $\theta = 3, T = 10, L = 20$

Fig. 1 shows some examples of low energy configurations obtained by MS on two-dimensional lattices. Note that the cavity approximation is not expected to be asymptotically exact on lattices, but that the configurations found have reasonably low energy.

COMPARISON WITH OTHER METHODS

We tested on random graphs two alternative strategies to solve the optimization problem of minimizing the energy $\mathscr{E}$: a Linear/Integer Programming (L/IP) approach and Simulated Annealing (SA). We tried two different L/IP formulations: one closely related to our representation based on x_i^t and the same dynamical constraints Ψ_i and a second one based on [15], solving both with the CPLEX software. In both cases performances were extremely poor (becoming unfeasible for graphs over a few

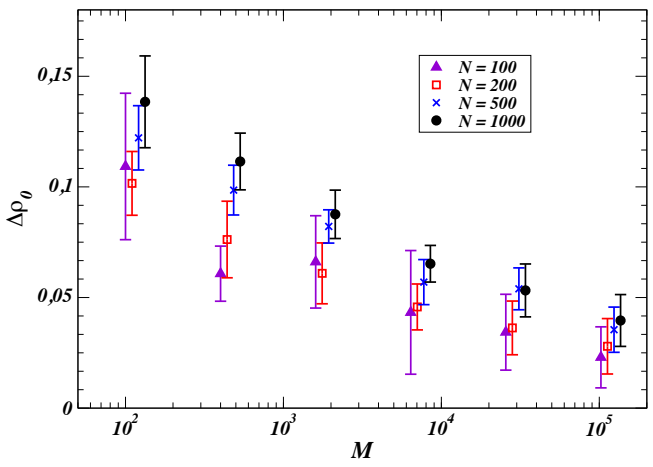

FIG. 2. Comparison between the performances of SA and MS algorithms on random regular graphs of degree $K = 5$, thresholds $\theta = 4$ and different sizes N. We computed the quantity $\Delta\rho_0 = \rho_0^{SA}/\rho_0^{MS} - 1$ (averaged over 10 realization of the graph) that represents the rescaled difference between the minimum density of seeds ρ_0^{SA} required to activate the whole graph obtained using SA with exponential cooling scheme from $\beta = 0.5$ to $\beta = 10^3$ and the corresponding values ρ_0^{MS} computed using the MS algorithm. The results for different system sizes N are plot as function of the total number of global updates M employed in the SA. Points are slightly shifted horizontally from the correct value at $N = 100$ for clarity. The running time of SA for each instance with $N = 1\,000, M = 102\,400$ was around 73 hours. For comparison, the same instances were solved by MS in less than 1 minute on the same machine.

dozen nodes) and will not be reported in detail. With SA we find that the time to reach the MS solution scales poorly with the size (see Figure 2) when the effective configurational space is sufficiently large. A detailed investigation of the performances of the algorithms depending on the choice of parameters (costs, weights, etc.) is beyond the scope of this Letter.

COLLECTIVE DYNAMICS UNDER OPTIMIZATION

For random initial conditions (i.e. i.i.d. seeds probability), there exists a density of seeds $\rho_0^c \leq 1$ above which the whole graph activates during the dynamics. Depending on the weights and the thresholds, the complete activation can occur continuously in ρ_0 or abruptly, as a result of some macroscopic cooperative mechanism. In the following, we show that the effects of the optimization are very different in case of continuous or discontinuous activation processes. For the sake of simplicity we consider instances of regular random graphs of degree K, with uniform weights $w_{ij} = 1$, $\forall(i,j)$, uniform thresholds $\theta_i = \theta$, $\forall i$ with $\theta \in \{1, \ldots, K-1\}$, and uniform costs $c_i = 1$, $\forall i$ (an especially difficult case for BP/MS due to the high degeneracy). The dynamics of the LTM is now reminiscent of bootstrap or k-core percolation processes [5, 6]. For random initial conditions, the collective behavior of the system depends on the values of θ. For $1 < \theta < K - 1$, the whole graph activates ($\rho_T = 1$) abruptly at a finite density ρ_0^c, whereas for $\theta = 1$ and $\theta = K - 1$ there is a continuous percolation-like transition.

We studied the ensemble of solutions of the problem in the single link approximation using population dynamics. In the completely homogeneous case (with $c_i = 1, \forall i$), the population simplifies to a single message. In general, MS and BP results prove that the optimization is effective when the activation process occurs continuously at $\epsilon = 0$. In particular for $\epsilon > 0$ a finite interval of values of $0 < \mu < \mu_c(\epsilon)$ exists in which full-spread solutions are achievable at $\rho_0 < \rho_0^c$. On the other hand, when the activation is discontinuous at $\epsilon = 0$, the single-link equations do not converge in the region of $\mu < \mu_c(\epsilon)$ where one can expect an improvement in the value of ρ_0 required for full activation.

Fig.3A displays the curves ρ_T vs. ρ_0 for $\theta = 2$ and $K = 3$ with and without optimization. The results obtained in the absence of optimization ($\epsilon = 0$) correspond to the solution of the dynamics starting from random initial conditions in which $x_i^0 = 1$ with probability ρ_0. While the activation is continuous for $\epsilon = 0$, increasing $\epsilon > 0$ we observe the emergence of a gap with the upper branch that optimizes ρ_T. The gap is just a manifestation of a first-order transition that is evident in Fig.3B, where we plot the curves $\rho_0(\mu)$ for $\epsilon = 0$ (black line) and 0.4 (red dashed line). Following the two branches of solutions across the transition point we find a coexistence region in which the upper branch becomes metastable (see blue lines in Fig.3B) and hysteresis phenomena can be observed. The convergence of the BP equations slows down in the optimized region where convergence time grows very fast by increasing μ along the upper branch. On the contrary, for decreasing μ in the lower branch the convergence is fast until the transition point where the equations do not converge anymore. The convergence problems are amplified by both increasing the inverse temperature β and/or the duration time T and can be reduced using a damped update rule or (large) populations of messages.

In Fig.4 we report the activation time statistics of the dynamical process. For random initial conditions, the activation probability $P(t)$ of a node decays exponentially with the time t as it usually occurs in spreading processes. Interestingly, in the region in which optimization is effective, the spreading process proceeds at a slower pace. For $\epsilon > 0$ and increasing μ from negative to positive values, the dynamics slows down and $P(t)$ develops a power-law behavior that persists until $\mu \simeq \mu_c$.

The results seem different in the case $\theta = 2$ and $K = 4$, in which the complete activation occurs abruptly even

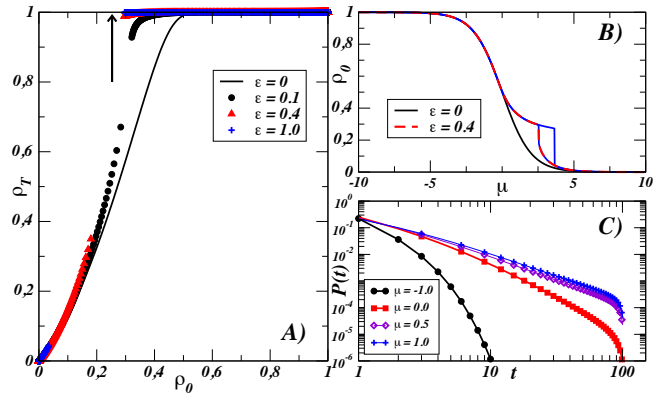

FIG. 3. A) Parametric plot ρ_T v.s.ρ_0 obtained solving Eqs. (2) in the single-link approximation on regular random graphs of degree $K = 3$, for threshold $\theta = 2$, duration $T = 20$ and $\epsilon = 0, 0.1, 0.4, 1$. The vertical arrow indicates the minimum density of seeds ($\rho_0 \approx 0.253$) necessary for the total activation obtained by MS on finite graphs of size $|V| = 10,000$. B) Curves $\rho_0(\mu)$ for $\epsilon = 0$ (black full line) and 0.4 (red dashed line). The blue curves are obtained following the upper and lower branches of solution across the transition. C) Activation time probability $P(t)$ obtained computing the total BP marginals in a dynamics of duration $T = 100$, for $\epsilon = 0.4$ and different values of μ.

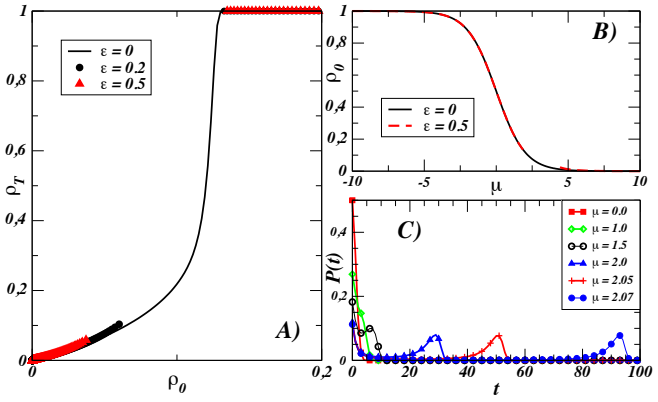

FIG. 4. Node activation probability $P(t)$ obtained computing the total BP marginals in the single-link approximation on regular random graphs of degree $K = 3$ (top) and $K = 4$ (bottom), threshold $\theta = 2$ with $T = 100$, $\epsilon = 0.5$ and different values of μ.

for $\epsilon = 0$. As reported in Fig.4A, the region of complete activation does not widen when increasing ϵ from zero, while the onset of discontinuity moves towards smaller values of ρ_0 and the BP equations do not converge in a finite interval of μ values (see Fig.4B). The analysis of the activation statistics for $\epsilon = 0$ reveals important information on the dynamics close to the abrupt transition. In Fig.4C we plot $P(t)$ for μ increasing in the optimized region: the non-monotonicity of $P(t)$ for $\mu > 0$ means that after a period of time in which the dynamics is slow and correlations are built throughout the graph, there is a sudden activation of a finite fraction of nodes that is reflected into the appearance of a bump in $P(t)$ at large times (that diverge for $\mu \to \mu_c$).

The above results show that, in case of continuous activation processes, some trajectories can be selected in order to optimize the spread of the process. But how rare are the optimal trajectories? The answer is given in Fig.5 (top) where we plot the entropy s of the initial conditions that lead to a full spread. This quantity is compared with the entropy associated to a random choice of initial conditions with fixed density of active nodes ρ_0 (binomial sampling without optimization). When the curves deviate from the binomial the probability of choosing randomly an optimal set of seeds becomes exponentially small (inset in Fig.5). In the infinite system, this event is governed by a zero-one law. The results for $N = 50, 100$ are obtained using a generalization of the cavity method that allows to fix a global constraint (the number of seeds) by introducing an additional set of messages that flow over a spanning tree superimposed on the original graph (see [13] Appendix B), the one for $N = 30$ by explicit enumeration.

RELATIONS TO EXISTING STUDIES

It is now clear that irreversible processes on graphs can be optimized using BP messages that depend on pairs of activation times for neighboring nodes. However, recent works concerning the zero-temperature dynamics of the random field Ising model [7] and the susceptible-infected model of epidemic spreading [8] suggest that in the absence of optimization, when the seeds are randomly drawn, the dynamics can be correctly analyzed using only single-time cavity marginals. In fact, for $\epsilon = 0$, the two-times formalism can be reduced to a set of cavity equations for $\chi_{i\ell}(t) = Pr_{-\ell}\{x_i^t = 1\}$, i.e. the probability that node i is active at time t in the absence of the neighbor ℓ. To show this, it is convenient to introduce a new representation of the dynamic rule. For each directed edge (i, ℓ), the variable $t_{i\ell}$ represents the time at which node i would activate in the absence of node ℓ, and for $i \notin S$, it satisfies the iterative equation $t_{i\ell} = \min_{\sum_{j\in\partial i\setminus\ell} w_{ji}\mathbb{1}[t_{ji}<t]\geq\theta_i} t$. As it happened for the activation times $\{t_i\}$, also the equations for the cavity-times admit a unique solution for a given choice of S, which is in one-to-one correspondence with the solution of the single t_i model. In order to optimize the trajectories and average over the initial conditions, we introduce the messages $\hat{H}_{i\ell}(t_{i\ell}, t_{\ell i})$ that satisfy a set of BP equations (see SM). When $\epsilon = 0$, the hypothesis that the messages $\{\hat{H}_{i\ell}(t_{i\ell}, t_{\ell i})\}$ do not depend on the backward cavity times $\{t_{\ell i}\}$ is self-consistently satisfied, and the BP equations can be easily reduced to single-time quantities $\hat{H}_{i\ell}(t_{i\ell})$. Integrating the resulting equations for $t_{i\ell} \leq t$ and averaging the contribution of

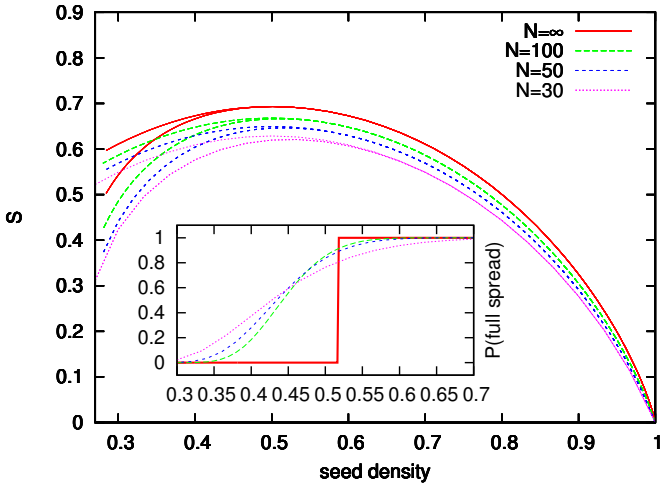

FIG. 5. Entropy s of the solutions of the full spread problem on regular random graphs of degree $K = 3$ and threshold $\theta = 2$, for $N = 30$, 50, 100, ∞ vs. seed density ρ_0. For each N the upper line corresponds to the binomial distribution (i.e. entropy of seeds in the absence of optimization) and the lower one to the entropy of fully spreading seeds. Inset: Probability P of randomly selecting a fully spreading set of seeds for the same set of parameters, i.e. the ratio of the corresponding two curves in the main plot.

all nodes, it is straightforward to obtain an evolution equation for the density $\rho(t)$ of active nodes at time t. Although the details of the model are different, the final single-time equations are equivalent to those derived in Ref.[7] (see also Ref.[8]). We remark that the two-times joint probability in Eq.2 is more than a technical artifact; on the contrary, it is quite crucial to allow information to flow backwards in time when optimizing over the final state ($\epsilon > 0$).

It is worth noting that our formulation in terms of activation times $\mathbf{t}$ is a direct simplification, seemingly valid only for microscopically irreversible processes, of the general dynamic cavity formulation recently proposed to study non-equilibrium dynamics on sparse graphs [10–12] (see SM). It could be in principle possible to relax the assumption of complete irreversibility, using probability distributions over time-dependent paths, but the optimization of fully reversible dynamics is currently numerically unfeasible and it requires new separate ideas to overcome current computational limitations.

Conclusions. – The study of inverse dynamical problems on large graphs is a new frontier for the application of message-passing algorithms and the approach presented here is directly applicable to a series of real-world problems. Nevertheless, our methods could be improved in two main directions. First, by including the optimization over some stochastic parameter [9], providing a formulation that is closer to a class of stochastic optimization problems that received a lot of attention in the computer science community [2].

The authors acknowledge the european grants FET Open 265496 and ERC 267915.

[1] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, *Nature* **473**, 167173 (2011).
[2] D. Kempe, J. Kleinberg, and E. Tardos, Proc. 9th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining (2003).
[3] J. Leskovec , L. A. Adamic , B. A. Huberman, Proc. 7th ACM conf. on Electronic Commerce, p.228-237, (2006).
[4] A. G. Haldane and R. M. May, *Nature* **469**, 351355 (2011).
[5] J. Chalupa, P.L. Leath, and G.R. Reich, *J. Phys. C: Solid State Phys.* **12** L31 (1979).
[6] S. N. Dorogovtsev, A. V. Goltsev, and J.F.F. Mendes, *Phys. Rev. Lett.* **96**, 040601 (2006).
[7] H. Ohta and S. Sasa, *Europhys. Lett.* **90**, 27008-1-5 (2010).
[8] B. Karrer and M. E. J. Newman, *Phys. Rev. E* **82**, 016101 (2010).
[9] F. Altarelli, A. Braunstein, A. Ramezanpour, and R. Zecchina, *Phys. Rev. Lett.* **106**, 190601 (2011).
[10] Y. Kanoria and A. Montanari, Ann. App. Prob., **21**(5), 2011.
[11] I. Neri and D. Bollé, *J. Stat. Mech.* (2009) P08009.
[12] E. Aurell and H. Mahmoudi, *J. Stat. Mech.* P04014 (2011); *Commun. Theor. Phys.* **56**, 157-162 (2011).
[13] A. Braunstein, A. Ramezanpour, R. Zecchina, and P. Zhang *Phys. Rev. E* **83**, 056114 (2011).
[14] Z. Lu, W. Zhang, W. Wu, J. Kim and B. Fu, *J. Comb. Optim.*, doi:10.1007/s10878-011-9393-3 (2011)
[15] E. Ackerman, O. Ben-Zwi and G. Wolfovitz, *Theor. Comp. Sci.* **411**, 44-46 (2010).

SUPPLEMENTAL MATERIAL

Derivation of the Max-Sum equations

The Max-Sum (MS) messages $h_{i\ell}(t_i, t_\ell)$ are defined in terms of the Belief Propagation messages $H_{i\ell}(t_i, t_\ell)$ as

$$h_{i\ell}(t_i, t_\ell) = \lim_{\beta\to\infty} \frac{1}{\beta} \log H_{i\ell}(t_i, t_\ell) + C_{i\ell} \tag{3}$$

where the additive constant $C_{i\ell}$ is such that

$$\max_{t_i, t_\ell} h_{i\ell}(t_i, t_\ell) = 0\,. \tag{4}$$

Taking the $\beta \to \infty$ limit of the BP equation (2) in the article we obtain

$$h_{i\ell}(t_i, t_\ell) = \begin{cases} \displaystyle\max_{\{t_j,\, j\in\partial i\setminus\ell\}} \left[\sum_{j\in\partial i\setminus\ell} h_{ji}(t_j, 0)\right] - \mu c_i + C_{i\ell} & \text{if } t_i = 0, \quad (5a) \\ \displaystyle\max_{\substack{\{t_j,\, j\in\partial i\setminus\ell\} \text{ s.t:} \\ \sum_{k\in\partial i} w_{ki}\mathbb{1}[t_k\le t_i-1]\ge\theta_i\,, \\ \sum_{k\in\partial i} w_{ki}\mathbb{1}[t_k<t_i-1]<\theta_i}} \left[\sum_{j\in\partial i\setminus\ell} h_{ji}(t_j, t_i)\right] + C_{i\ell} & \text{if } 0 < t_i \le T, \quad (5b) \\ \displaystyle\max_{\substack{\{t_j,\, j\in\partial i\setminus\ell\} \text{ s.t:} \\ \sum_{k\in\partial i} w_{ki}\mathbb{1}[t_k<T]<\theta_i}} \left[\sum_{j\in\partial i\setminus\ell} h_{ji}(t_j, \infty)\right] - 1 + C_{i\ell} & \text{if } t_i = \infty. \quad (5c) \end{cases}$$

These equations are derived in a straightforward way, but their computational complexity is exponential in the connectivity of the node being considered. We shall now see how to reduce them to a form which can be computed efficiently. The same method can be easily extended to the BP equations.

Simplification of the update equations

The maximization in (5a) is unconstrained, and it reduces to

$$h_{i\ell}(0, t_\ell) = \sum_{j\in\partial i\setminus\ell} \left[\max_{t_j} h_{ji}(t_j, 0)\right] - \mu c_i + C_{ij} \tag{6}$$

which is trivial to compute.

The constrained maximisations in (5b, 5c) can be computed efficiently by iteration. To compute (5b) we introduce (with an obvious simplification of notation):

$$Q_{1,\dots,r}(\theta_1, \theta_2, t) = \max_{\substack{\{t_1,\dots,t_r\} \text{ s.t:} \\ \sum_{j=1,\dots,r} w_j\mathbb{1}[t_j\le t-1]=\theta_1\,, \\ \sum_{j=1,\dots,r} w_j\mathbb{1}[t_j<t-1]=\theta_2}} \left[\sum_{j=1,\dots,r} h_j(t_j, t)\right] \tag{7}$$

$$= \max_{t_r} \left\{h_r(t_r, t) + Q_{1,\dots,r-1}\big(\theta_1 - w_r\mathbb{1}\left[t_r \le t-1\right], \theta_2 - w_r\mathbb{1}\left[t_r < t-1\right], t\big)\right\} \tag{8}$$

with the first element directly obtained from the definition:

$$Q_1(\theta_1, \theta_2, t) = \max_{\substack{t_1 \text{ s.t:} \\ w_1\mathbb{1}[t_1\le t-1]=\theta_1 \\ w_1\mathbb{1}[t_1<t-1]=\theta_2}} h_1(t_1, t) \tag{9}$$

$$= \begin{cases} \max_{t_1>t-1} h_1(t_1, t) & \text{if } \theta_1 = \theta_2 = 0, \\ h_1(t-1, t) & \text{if } \theta_1 = w_1 \text{ and } \theta_2 = 0, \\ \max_{t_1<t-1} h_1(t_1, t) & \text{if } \theta_1 = \theta_2 = w_1, \\ -\infty & \text{otherwise.} \end{cases} \tag{10}$$

The convolution of two Q's is given by

$$Q_{1,\dots,r}(\theta_1,\theta_2,t) = \max_{\substack{\theta'_1,\theta''_1,\theta'_2,\theta''_2 \text{ s.t:}\\ \theta'_1+\theta''_1=\theta_1,\,\theta'_2+\theta''_2=\theta_2}} \left\{Q_{1,\dots,s}(\theta'_1,\theta'_2,t) + Q_{s+1,\dots,r}(\theta''_1,\theta''_2,t)\right\} . \tag{11}$$

Once $Q(\theta_1,\theta_2,t)$ is computed from the convolution of all the incoming edges, we can easily compute

$$M(\theta_1,\theta_2,t) = \max_{\substack{\{t_1,\dots,t_r\} \text{ s.t:}\\ \sum_{j=1,\dots,r} w_j \mathbb{1}[t_j\le t-1]\ge\theta_1\,,\\ \sum_{j=1,\dots,r} w_j \mathbb{1}[t_j< t-1]<\theta_2}} \left\{\sum_{j=1,\dots,r} H_j(t_j,t)\right\} \tag{12}$$

$$= \max_{\substack{\{\theta'_1,\theta'_2\} \text{ s.t:}\\ \theta'_1\ge\theta_1,\theta'_2<\theta_2}} Q(\theta'_1,\theta'_2,t) \tag{13}$$

in terms of which (5b) gives:

$$h_{i\ell}(t_i,t_\ell) = M\big(\theta_i - w_{\ell i}\mathbb{1}\left[t_\ell \le t_i - 1\right], \theta_i - w_{\ell i}\mathbb{1}\left[t_\ell < t_i - 1\right], t_i\big) + C_{i\ell} \qquad \text{if } 0 < t_i \le T\,. \tag{14}$$

To compute (5c) we introduce

$$S_{1,\dots,r}(\theta) = \max_{\substack{\{t_1,\dots,t_r\} \text{ s.t:}\\ \sum_{j=1,\dots,r} w_j \mathbb{1}[t_j<T]=\theta}} \left[\sum_{j=1,\dots,r} h_j(t_j,\infty)\right] \tag{15}$$

$$= \max_{t_r} \left\{h_r(t_r,d) + S_{1,\dots,r-1}\big(\theta - w_r\mathbb{1}\left[t_r < T\right]\big)\right\} \tag{16}$$

with first element

$$S_1(\theta) = \begin{cases} \max\left\{h_1(T,\infty), h_1(\infty,\infty)\right\} & \text{if } \theta = 0\\ \max\limits_{t_1<T} h_1(t_1,\infty) & \text{if } \theta = w_1\\ -\infty & \text{otherwise}\end{cases} \tag{17}$$

and generic convolution

$$S_{1,\dots,r}(\theta) = \max_{\theta',\theta'' \text{ s.t: } \theta'+\theta''=\theta} \left\{S_{1,\dots,s}(\theta') + S_{s+1,\dots,r}(\theta'')\right\} . \tag{18}$$

Once S is computed for all the incoming edges, we can compute

$$P(\theta) = \max_{\substack{\{t_1,\dots,t_r\} \text{ s.t:}\\ \sum_{j=1,\dots,r} w_j \mathbb{1}[t_j<T]<\theta}} \left\{\sum_{j=1,\dots,r} h_j(t_j,\infty)\right\} \tag{19}$$

$$= \max_{\theta' \text{ s.t: } \theta'<\theta} S(\theta') \tag{20}$$

in terms of which (5c) becomes:

$$h_{i\ell}(\infty,t_\ell) = P\big(\theta_i - w_{\ell i}\mathbb{1}\left[t_\ell < T\right]\big) - 1 + C_{i\ell}. \tag{21}$$

Notice that $P(\theta) = M(0,\theta,t)$ with $t=\infty$, so we can compute it in the same manner as $M(\theta_1,\theta_2,t)$ provided we extend the range of values for t so as to include $t=\infty$.

Simplification of the messages

The update equations can be further simplified by noticing that in (5) the dependence of $h_{i\ell}$ on t_ℓ is almost trivial: for fixed t_i, $h_{i\ell}$ is independent on t_ℓ if $t_i = 0$, it only depends on sign $\big(t_\ell - (t_i-1)\big) \in \{-1,0,1\}$ if $0 < t_i \le T$, and it

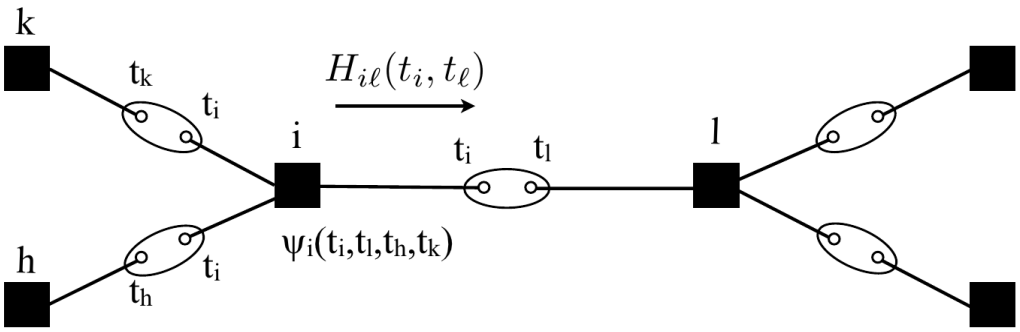

FIG. 6. Dual factor graph representation for the Spread problem

only depends on $\mathbb{1}[t_\ell < T]$ if $t_i = \infty$. We then introduce:

$$\tilde{h}_{i\ell}(t_i, \sigma) = \begin{cases} h_{i\ell}(t_i, t_i - 2) & \text{if } t_i > 1 \text{ and } \sigma = 0 \\ h_{i\ell}(t_i, t_i - 1) & \text{if } t_i > 0 \text{ and } \sigma = 1 \\ h_{i\ell}(t_i, t_i) & \text{if } \sigma = 2 \\ -\infty & \text{otherwise} \end{cases} \tag{22}$$

which can be inverted as

$$h_{i\ell}(t_i, t_\ell) = \tilde{h}_{i\ell}\big(t_i, \hat{\sigma}(t_i, t_\ell)\big) \tag{23}$$

$$\hat{\sigma}(t_i, t_\ell) = 1 + \text{sign}\left(t_\ell - (t_i - 1)\right) \tag{24}$$

and in terms of which (8), (10), (16) and (17) become respectively:

$$Q_{1,\dots,r}(\theta_1, \theta_2, t) = \max_{t_r} \left\{ \tilde{h}_r\big(t_r, \hat{\sigma}(t_r, t)\big) + Q_{1,\dots,r-1}\big(\theta_1 - w_r \mathbb{1}[t_r \le t-1], \theta_2 - w_r \mathbb{1}[t_r < t-1], t\big) \right\} \tag{25}$$

$$Q_1(\theta_1, \theta_2, t) = \begin{cases} \max\limits_{t_1 > t-1} \tilde{h}_1\big(t_1, \hat{\sigma}(t_1, t)\big) & \text{if } \theta_1 = \theta_2 = 0 \\ \tilde{h}_1(t-1, 2) & \text{if } \theta_1 = w_1 \text{ and } \theta_2 = 0 \\ \max\limits_{t_1 < t-1} \tilde{h}_1(t_1, 2) & \text{if } \theta_1 = \theta_2 = w_1 \\ -\infty & \text{otherwise} \end{cases} \tag{26}$$

$$S_{1,\dots,r}(\theta) = \max_{t_r} \left\{ \tilde{h}_r(t_r, 2) + S_{1,\dots,r-1}\big(\theta - w_r \mathbb{1}[t_r < T]\big) \right\} \tag{27}$$

$$S_1(\theta) = \begin{cases} \max\left\{ \tilde{h}_1(d-1, 2), \tilde{H}_1(d, 2) \right\} & \text{if } \theta = 0 \\ \max\limits_{t_1 < d-1} \tilde{h}_1(t_1, 2) & \text{if } \theta = w_1 \\ -\infty & \text{otherwise} \end{cases} \tag{28}$$

and (6), (14) and (21) become:

$$\tilde{h}_{i\ell}(t_i, \sigma) = \begin{cases} \sum\limits_{j \in \partial i \setminus \ell} \left[\max\limits_{t_j} \tilde{h}_{ji}\big(t_j, \hat{\sigma}(t_j, 0)\big) \right] - \mu c_i + C_{i\ell} & \text{if } t_i = 0 \text{ and } \sigma = 2, & \text{(29a)} \\ M(\theta_i - w_{\ell i}, \theta_i - w_{\ell i}, t_i) + C_{i\ell} & \text{if } 1 < t_i \le T \text{ and } \sigma = 0, & \text{(29b)} \\ M(\theta_i - w_{\ell i}, \theta_i, t_i) + C_{i\ell} & \text{if } 0 < t_i \le T \text{ and } \sigma = 1, & \text{(29c)} \\ M(\theta_i, \theta_i, t_i) + C_{i\ell} & \text{if } 0 < t_i \le T \text{ and } \sigma = 2, & \text{(29d)} \\ P(\theta_i - w_{\ell i}) - 1 + C_{i\ell} & \text{if } t_i = \infty \text{ and } \sigma = 0, & \text{(29e)} \\ P(\theta_i) - 1 + C_{i\ell} & \text{if } t_i = \infty \text{ and } \sigma > 0, & \text{(29f)} \\ -\infty & \text{otherwise.} & \text{(29g)} \end{cases}$$

Time complexity of the updates

A naive implementation of 29a-29g would require $O(Td(d-1)\Theta^4)$ operations for a vertex of degree d, but this can be reduced substantially to $O(Td\Theta^2)$ as follows. A factor $d-1$ can be saved by pre-computing the convolution $Q_{1,\dots,i}$ and $Q_{i,\dots,d}$ for each $i = 1, ..., d$ using $2d$ convolution operations, and then computing "cavity" $Q_{1,\dots,i-1,i+1,\dots d}$ as a convolution of $Q_{1,\dots,i-1}$ and $Q_{i+1,\dots,d}$. Moreover, a factor Θ^2 can be saved by employing FFT to compute convolutions in the case of BP.

For the case of MS, the problem can be relaxed as follows in order to reduce complexity. We consider the relaxed constraint $\Psi'_i = \mathbb{1}[t_i = 0] + \mathbb{1}[t_i \geq \phi(\{t_j\}_{j\in i})]$ instead of Ψ_i. The formal consequence is that the equivalent of Eq. 5 has only one inequality, and thus the complexity of the max-sum convolution becomes $O(Td\Theta^2)$. Notice that in the case of $\epsilon = 0$, this is equivalent to considering the following dynamics: at each time t, select a subset of the nodes and update them using the standard dynamical rule. In the optimization case $\epsilon > 0$ unfortunately this is no longer true.

BP equations on Regular Random Graphs

We consider here the ensemble of random regular graphs of degree K on which we define for simplicity a completely homogeneous LTM with uniform weights $w_{ij} = 1\ \forall (i,j) \in E$, uniform thresholds $\theta_i = \theta\ \forall i \in V$, and uniform costs $c_i = 1\ \forall i \in V$. Because of the homogeneity of the problem, in the infinite size limit ($N \to \infty$) one can consider a single-link approximation, obtaining the following system of nonlinear equations:

$$
\begin{aligned}
H(t,s) &\propto \sum_{\substack{n_-+n_++n_0=K-1 \\ n_-<\theta-\mathbb{1}[s<t-1] \\ \theta-\mathbb{1}[s\leq t-1]\leq n_-+n_0}} \frac{K-1!}{n_-!n_+!n_0!} p_t^{K-1-n_--n_0} m_t^{n_-} H(t-1,t)^{n_0} \qquad \text{for} \quad 0 < t \leq T \\
H(0,s) &\propto e^{-\beta\mu} p_0^{K-1} \\
H(\infty,s) &\propto e^{-\beta\epsilon} \sum_{n_-\leq\theta-1-\mathbb{1}[s<T]} \binom{K-1}{n_-} \left[H(T,\infty)+H(\infty,\infty)\right]^{K-1-n_-} m_\infty^{n_-}
\end{aligned}
\tag{30}
$$

where we defined the cumulative messages $p_t = \sum_{s\geq t} H(s,t)$ and $m_t = \sum_{s<t-1} H(s,t)$. The normalization constant is just the sum of all messages. The system of equations could be further simplified from $O(T^2)$ messages to $O(T)$ by exploiting the fact that $H(t,s) = H(t, \text{sign}(t-s+1))$. As this reduction requires the introduction of slightly more complex normalization constant, for simplicity here we do not consider it explicitly.

In general one should study the behavior of Eqs.30 varying μ, ϵ, β and T for any given assignment of K and θ. Here we consider only two exemplar cases: 1) $K = 3, \theta = 2$, and 2) $K = 4, \theta = 2$. The behavior in the (ϵ, μ)-plane is studied at fixed $T = 20$ and $\beta = 1$.

BP equations in the cavity-times representation

In the cavity-times representation of the dynamics, a variable $t_{i\ell}$ represents the time at which a node i would activate in the absence of node ℓ. Let us define:

$$
f_i\left(\{t_k\}_{k\in\partial i}\right) = \min\left\{t : \sum_{k\in\partial i} w_{ki} 1\left[t_k < t\right] \geq \theta_i\right\} \tag{31}
$$

$$
f_{ij}\left(\{t_k\}_{k\in\partial i\setminus j}\right) = \min\left\{t : \sum_{k\in\partial i\setminus j} w_{ki} 1\left[t_k < t\right] \geq \theta_i\right\} \tag{32}
$$

when i is not a seed, $t_{i\ell}$ satisfies the iterative equations $t_{i\ell} = f_{i\ell}\left(\{t_k\}_{k\in\partial i\setminus\ell}\right)$. The messages $\hat{H}_{i\ell}\left(t_{i\ell}, t_{\ell i}\right)$ satisfy the following BP equations

$$\begin{aligned}\hat{H}_{i\ell}\left(t_{i\ell}, t_{\ell i}\right) &\propto \sum_{\{t_{ki}, t_{ik}\}_{k\in\partial i\setminus\ell}} \prod_{k\in\partial i\setminus\ell} \hat{H}_{ki}\left(t_{ki}, t_{ik}\right) \left\{ \prod_{k\in\partial i} \delta\left(t_{ik}, f_{i\ell}\left(\{t_{k'i}\}_{k'\in i\setminus k}\right)\right) e^{-\epsilon\delta\left(f_i\left(\{t_{ki}\}_{k\in\partial i}\right),\infty\right)} + \prod_{k\in\partial i} \delta\left(t_{ik}, 0\right) e^{-\mu} \right\} \\ &= \sum_{\{t_{ki}\}_{k\in i\setminus\ell}} \delta\left(t_{i\ell}, f_{i\ell}\left(\{t_{ki}\}_{k\in\partial i\setminus\ell}\right)\right) e^{-\epsilon\delta\left(f_{i\ell}\left(\{t_{ki}\}_{k\in\partial i}\right),\infty\right)} \prod_{k\in\partial i\setminus\ell} \hat{H}_{ki}\left(t_{ki}, f_{i\ell}\left(\{t_{k'i}\}_{k'\in\partial i\setminus k}\right)\right) + \\ &\quad + \delta\left(t_{i\ell}, 0\right) \prod_{k\in\partial i\setminus\ell} \hat{H}_{ki}\left(t_{ki}, 0\right) e^{-\mu}.\end{aligned}$$

Assuming $\hat{H}_{k\to i}\left(t_{ki}, y\right) = \hat{H}_{k\to i}\left(t_{ki}\right)$ we get

$$\hat{H}_{i\ell}\left(t_{i\ell}, t_{\ell i}\right) \propto \sum_{\{t_{ki}\}_{k\in\partial i\setminus\ell}} \prod_{k\in\partial i\setminus\ell} \hat{H}_{ki}\left(t_{ki}\right) \left\{ \delta\left(t_{i\ell}, f_{i\ell}\left(\{t_{ki}\}_{k\in\partial i\setminus\ell}\right)\right) e^{-\epsilon\delta\left(f_i\left(\{t_{ki}\}_{k\in\partial i}\right),\infty\right)} + \delta\left(t_{i\ell}, 0\right) e^{-\mu} \right\} \tag{33}$$

We will show that in the case $\epsilon = 0$, the equation becomes equivalent to those derived for the zero-temperature random-field Ising model by Ohta and Sasa [7]. For that purpose, we need to consider time-cumulative quantities, such as the probability $\chi_{i\ell}(t) = \sum_{t_{i\ell}\le t} \hat{H}_{i\ell}(t_{i\ell})$ that node i is active at time t in the absence of node ℓ. Let us first define the following sequence of increasing sets U_t

$$\begin{aligned} V_t &= \left\{ \{t_{ki}\} : \sum_{k\in\partial i\setminus\ell} w_{ki} \mathbb{1}\left[t_{ki} < t\right] \ge \theta_i \wedge \sum_{k\in\partial i\setminus\ell} w_{ki} \mathbb{1}\left[t_{ki} < t-1\right] < \theta_i \right\} \\ U_t &= \left\{ \{t_{ki}\} : \sum_{k\in\partial i\setminus\ell} w_{ki} \mathbb{1}\left[t_{ki} < t\right] \ge \theta_i \right\} \\ U_{t+1} &= U_t \cup V_{t+1} \\ U_t \cap V_{t+1} &= \emptyset \\ \mathbb{1}\left[U_t\right] &= \sum_{0<t'\le t} \mathbb{1}\left[V_{t'}\right] \\ &= \sum_{0<t'\le t} \delta\left(t', f_{i\ell}(\{t_{ki}\}_{k\in\partial i\setminus\ell})\right) \end{aligned}$$

With this definitions, let us analyze the evolution equations Eq.33 in the $\epsilon = 0$ case. It is easy to see that under the hypothesis that all $H_{ki}(t_{ki}, t_{ik})$ on the rhs. do not depend on t_{ki} and $\epsilon = 0$, then the dependence on $t_{\ell i}$ disapears. This implies that the hypothesis is self-consistent and will be verified at every iteration if it is verified at the initial one (e.g. if the messages have uniform initialization). Now Eq. 33 becomes

$$\hat{H}_{i\ell}\left(t_{i\ell}\right) \propto \sum_{\{t_{ki}\}_{k\in\partial i\setminus\ell}} \left\{ \delta\left(t_{i\ell}, f_{i\ell}\left(\{t_{ki}\}_{k\in\partial i\setminus\ell}\right)\right) + \delta\left(t_{i\ell}, 0\right) e^{-\mu} \right\} \prod_{k\in\partial i\setminus\ell} \hat{H}_{ki}\left(t_{ki}\right)$$

Let us compute the time-cumulative quantities:

$$
\begin{aligned}
\chi_{i\ell}(t) &= \sum_{t_{i\ell}\leq t} \hat{H}_{i\ell}\left(t_{i\ell}\right) \\
&\propto \sum_{\{t_{ki}\}_{k\in\partial i\setminus\ell}} \left\{ \sum_{0<t_{i\ell}\leq t} \delta\left(t_{i\ell}, f_{i\ell}\left(\{t_{ki}\}_{k\in\partial i\setminus\ell}\right)\right) + e^{-\mu} \right\} \prod_{k\in\partial i\setminus\ell} \hat{H}_{ki}\left(t_{ki}\right) \\
&= e^{-\mu} + \sum_{\{t_{ki}\}_{k\in\partial i\setminus\ell}} \mathbb{1}\left[\sum_{k\in\partial i\setminus\ell} w_{ki}\mathbb{1}\left[t_{ki}<t\right] \geq \theta_i \right] \prod_{k\in\partial i\setminus\ell} \hat{H}_{ki}\left(t_{ki}\right) \\
&= e^{-\mu} + \sum_{\{t_{ki}\}_{k\in\partial i\setminus\ell}} \sum_{\{x_k=0,1\}} \prod_{k\in\partial i\setminus\ell} \delta\left(x_k, \mathbb{1}\left[t_{ki}<t\right]\right) \mathbb{1}\left[\sum_{k\in\partial i\setminus\ell} w_{ki}\mathbb{1}\left[t_{ki}<t\right] \geq \theta_i \right] \prod_{k\in\partial i\setminus\ell} \hat{H}_{ki}\left(t_{ki}\right) \\
&= e^{-\mu} + \sum_{\{x_k=0,1\}} \mathbb{1}\left[\sum_{k\in\partial i\setminus\ell} w_{ki}x_k \geq \theta_i \right] \prod_{k\in\partial i\setminus\ell} \sum_{t_{ki}} \hat{H}_{ki}\left(t_{ki}\right) \delta\left(x_k, \mathbb{1}\left[t_{ki}<t\right]\right) \\
&= e^{-\mu} + \sum_{\{x_k=0,1\}} \mathbb{1}\left[\sum_{k\in\partial i\setminus\ell} w_{ki}x_k \geq \theta_i \right] \prod_{k\in\partial i\setminus\ell} \left\{ x_k \sum_{t_{ki}<t} \hat{H}_{ki}\left(t_{ki}\right) + (1-x_k) \sum_{t_{ki}\geq t} \hat{H}_{ki}\left(t_{ki}\right) \right\} \\
&= e^{-\mu} + \sum_{\{x_k=0,1\}} \mathbb{1}\left[\sum_{k\in\partial i\setminus\ell} w_{ki}x_k \geq \theta_i \right] \prod_{k\in\partial i\setminus\ell} \left\{ x_k\chi_{ki}\left(t-1\right) + (1-x_k\chi_{ki}\left(t-1\right)) \right\}
\end{aligned}
$$

Denoting by $\alpha_i = e^{-\mu}/(1+e^{-\mu})$ the probability of choosing i as a seed and the initial conditions are $\chi_{i\ell}(0) = \alpha_i \ \forall i$ and fixing the normalization factor $(1+e^{-\mu})^{-1} = 1-\alpha_i$, we obtain

$$
\chi_{i\ell}(t+1) = \alpha_i + (1-\alpha_i) \sum_{\{x_k=0,1\}} \mathbb{1}[\sum_{k\in\partial i\setminus\ell} x_k w_{ki} \geq \theta_i] \prod_{k\in\partial i\setminus\ell} [x_k\chi_{ki}(t) + (1-x_k)(1-\chi_{ki}(t))] \tag{34}
$$

From which we can compute finally the density of active nodes $\rho(t) = \frac{1}{N}\sum_i \rho_i(t)$. Apart from the obvious differences in the details of the dynamical update, the single-link version of Eq. 34 (assuming all χ_{ki} identical) is substantially equivalent to Eq.(4) in Ref.[7].

Relation to the Dynamic Cavity Equations

A general method to study non-equilibrium dynamical processes on sparse graphs has been recently introduced by several authors under the name of *dynamic cavity method* [10–12]. It is easy to show that in the absence of optimization ($\epsilon = 0$), the belief-propagation equations presented in the main text can be viewed as a simplified version, valid only for microscopically irreversible processes, of the dynamic cavity equations. To this end, we adopt a formulation similar to that used by Neri and Bollé [11] by considering the path probability $P(\underline{\mathbf{x}}^t|\underline{\mathbf{h}}^t)$ of a trajectory $\underline{\mathbf{x}}^t = \{\mathbf{x}^0, \ldots, \mathbf{x}^t\}$ with $\mathbf{x}^t = \{x_0^t, \ldots, x_N^t\}$ in the presence of an external field $\underline{\mathbf{h}}^t = \{\mathbf{h}^0, \ldots, \mathbf{h}^t\}$. On a cavity graph, in which node i and its interactions are removed, the path probability can be written as

$$
P(\underline{\mathbf{x}}^t|\underline{\mathbf{h}}^t) = P_{(i)}(\underline{\mathbf{x}}^t|\underline{\mathbf{h}}^t + \underline{\mathbf{u}}_{(i)}^t) \prod_{s=1}^{t} W\left[x_i^s|\mathbf{x}^{s-1}; h_i^s\right] p_0(x_i^0) \tag{35}
$$

where $\underline{\mathbf{u}}_{(i)}^t$ is an auxiliary external field acting over the neighbors of i and introduced to keep track of the directed influence of i over them along the dynamics. For simplicity we have taken a factorized distribution over the initial conditions. By summing Eq.35 over all possible trajectories of all nodes $j \neq i$, we get the local marginal $P_i(\underline{x}_i^t|\underline{\mathbf{h}}^t + \underline{\mathbf{u}}_{(i)}^t)$

that satisfies the equation

$$P_i(\underline{x}_i^t|\underline{\mathbf{h}}^t+\underline{\mathbf{u}}_{(i)}^t)=\sum_{\underline{\mathbf{y}}_{\partial i}}P_{(i)}(\underline{\mathbf{y}}_{\partial i}^t|\underline{\mathbf{h}}^t+\underline{\mathbf{u}}_{(i)}^t)\prod_{s=1}^{t}W\left[x_i^s|\mathbf{x}^{s-1};h_i^s\right]p_0(x_i^0)$$
$$=\sum_{\underline{\mathbf{y}}_{\partial i}}P_{(i)}(\underline{\mathbf{y}}_{\partial i}^{t-1}|\underline{\mathbf{h}}^{t-1}+\underline{\mathbf{u}}_{(i)}^{t-1})\prod_{s=1}^{t}W\left[x_i^s|\mathbf{x}^{s-1};h_i^s\right]p_0(x_i^0) \tag{36}$$

where $\mathbf{y}_{\partial i}^t$ is the joint trajectory of the neighbors of i up to time t, and the last passage comes from the fact that the dynamics is parallel and the transition probability for variable i at time t depends only on the states of neighbors at time $t-1$. Exploiting the tree-like assumption, we can express the path probability distribution over the neighbors of i in factorized form

$$P_{(i)}(\underline{\mathbf{y}}_{\partial i}^{t-1}|\underline{\mathbf{h}}^{t-1}+\underline{\mathbf{u}}_{(i)}^{t-1})\propto\prod_{j\in\partial i}P_{j\to i}(\underline{x}_j^{t-1}|\underline{h}_j^{t-1}+\underline{u}_{ji}^{t-1}) \tag{37}$$

In the present case, we have $h_i^s=\theta_i$ and $u_{ji}^s=w_{ij}x_i^{s-1}$ for all times $s>0$, while the transition probability is given by the deterministic update rule as follows

$$W\left[x_i^s=0|\mathbf{x}^{s-1};h_i^s\right]=\mathbb{1}\left[\sum_{k\in\partial i}w_{ki}x_k^{s-1}<\theta_i\right], \tag{38}$$

$$W\left[x_i^s=1|\mathbf{x}^{s-1};h_i^s\right]=\mathbb{1}\left[\sum_{k\in\partial i}w_{ki}x_k^{s-1}\geq\theta_i\right]. \tag{39}$$

For a given directed edge (i,j), thus the local field can be univocally parametrized in terms of the variable in the removed node, leading to the following set of recursive equations for the cavity marginals $P_{i\to\ell}(\underline{x}_i^t|\underline{h}_i^t+\underline{u}_{i\ell}^t)\equiv P_{i\to\ell}(\underline{x}_i^t,\underline{x}_\ell^t)$,

$$P_{i\to\ell}(\underline{x}_i^t,\underline{x}_\ell^t)\propto p_0(x_i^0)\sum_{\{\underline{x}_j^{t-1}\}_{j\in\partial i\setminus\ell}}\prod_{s=1}^{t}W\left[x_i^s|\mathbf{x}^{s-1};\theta_i\right]\prod_{j\in\partial i\setminus\ell}P_{j\to i}(\underline{x}_j^{t-1},\underline{x}_i^{t-1}) \tag{40}$$

A path probability is now a vector of $O(4^{t+1})$ variables, but further reduction is possible because the dynamics is microscopically irreversible. The full sequence of $t+1$ binary values taken by $\underline{x}_i^t=\{x_i^0,\dots,x_i^t\}$ can be encoded in a single integer $t_i=\{0,\dots,t\}$ representing the time at which the variable i flips from 0 to 1 (with the convention $t_i=0$ for a seed). With this parametrization, the dynamic cavity equations 40 take the form

$$P_{i\to\ell}(t_i,t_\ell)\propto p_0(\delta_{t_i,0})\sum_{\{t_j\}_{j\in\partial i\setminus\ell}}\mathbb{1}\left[\sum_{j\in\partial i}w_{ji}\mathbb{1}[t_j<t_i-1]<\theta_i\right]\mathbb{1}\left[\sum_{j\in\partial i}w_{ji}\mathbb{1}[t_j<t_i]\geq\theta_i\right]\prod_{j\in\partial i\setminus\ell}P_{j\to i}(t_j,t_i)$$
$$=e^{-\mu c_i\delta_{t_i,0}}\sum_{\{t_j\}_{j\in\partial i\setminus\ell}}\mathbb{1}\left[\sum_{j\in\partial i}w_{ji}\mathbb{1}[t_j<t_i-1]<\theta_i\right]\mathbb{1}\left[\sum_{j\in\partial i}w_{ji}\mathbb{1}[t_j<t_i]\geq\theta_i\right]\prod_{j\in\partial i\setminus\ell}P_{j\to i}(t_j,t_i) \tag{41}$$

where we represented the initial factorized distribution in terms of weights over the seeds, i.e. $p_0(\delta_{t_i,0})=e^{-\mu c_i\delta_{t_i,0}}/(1+e^{-\mu c_i\delta_{t_i,0}})$. It is easy to check that Eq.41 corresponds to the $\epsilon=0$ case of the BP equations derived in the main text for the cavity messages $H_{i\to\ell}(t_i,t_\ell)$. From the above derivation it is also evident that optimization could be included by introducing in the dynamic cavity equation a local energy term $\mathcal{E}=\sum_i\mathcal{E}_i(x_i)=-\sum_i\sum_{s\leq t}\epsilon\delta_{x_i^t,1}$ in order to account for optimization over the trajectories. In the activation time representation the local energy term correctly becomes $\mathcal{E}_i=\epsilon\delta_{t_i,\infty}$.